

Preston Smith

Director of Research
Services

May 18, 2016

HTCONDOR WEEK
2016

A photograph of the Purdue University archway, a large black metal structure with two white stone pillars, set against a clear blue sky. In the background, a large brick building with a curved roof is visible. The foreground is a dark, textured area with a diagonal line pattern.

RESEARCH DATA DEPOT
AT PURDUE UNIVERSITY



(Dramatization)

Ran into Miron at a workshop recently..

Talked about data and the challenges of providing that service to campus.

Miron: “I’d like to talk about that at HTCondor Week!”

Campuses and HTCondor sites all face challenges with storage, so..



COMMUNITY

CLUSTERS

**A BUSINESS MODEL FOR COMPUTING AT
PURDUE**

- You get out at least what you put in
 - Buy 1 node or 100, you get a queue that guarantees access up to that many CPUs
- But wait, there's more!!
 - What if your neighbor isn't using his queue?
 - You can use it, but your job is subject to a time limit
- You don't have to do the work
 - Your grad student gets to do research rather than run your cluster.
 - Nor do you have to provide space in your lab for computers.
 - IT provides data center space, systems administration, application support.
 - Just submit jobs!

- 5 Year cycle
 - We build a cluster **every** year!
 - Vendors provide 5 year warranty
 - After 5 years, MOU with faculty says that the cluster will be retired
 - Faculty get credit for the remaining market value of their nodes, towards the next cluster.
 - Community clusters now appear to funding agencies as paying for a service – not a capital purchase.

CHALLENGES

CHALLENGES TO SMALLER COMMUNITIES

- HPC and HTC communities prefer different points to optimize the scheduler.
- Small but key communities (like large memory) lose benefits of standby queues when fewer nodes are spread between several clusters.
- HTC or large memory communities often have little need for HPC-specific optimizations
 - Accelerators
 - High-speed, low-latency networks

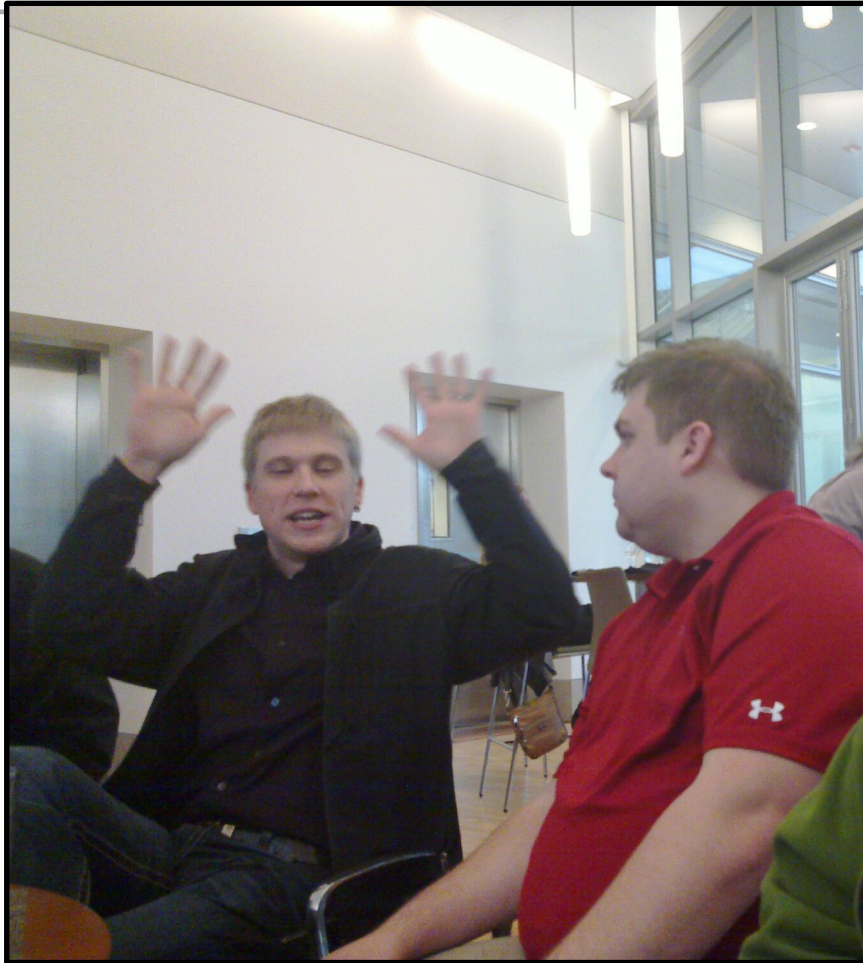
Emerging communities often don't fit in existing model at all!

Big Data Analytics

Graphics Rendering

Nontraditional platforms (Windows, cloud)

OBLIGATORY HTC/HTCONDOR CONTENT COMING



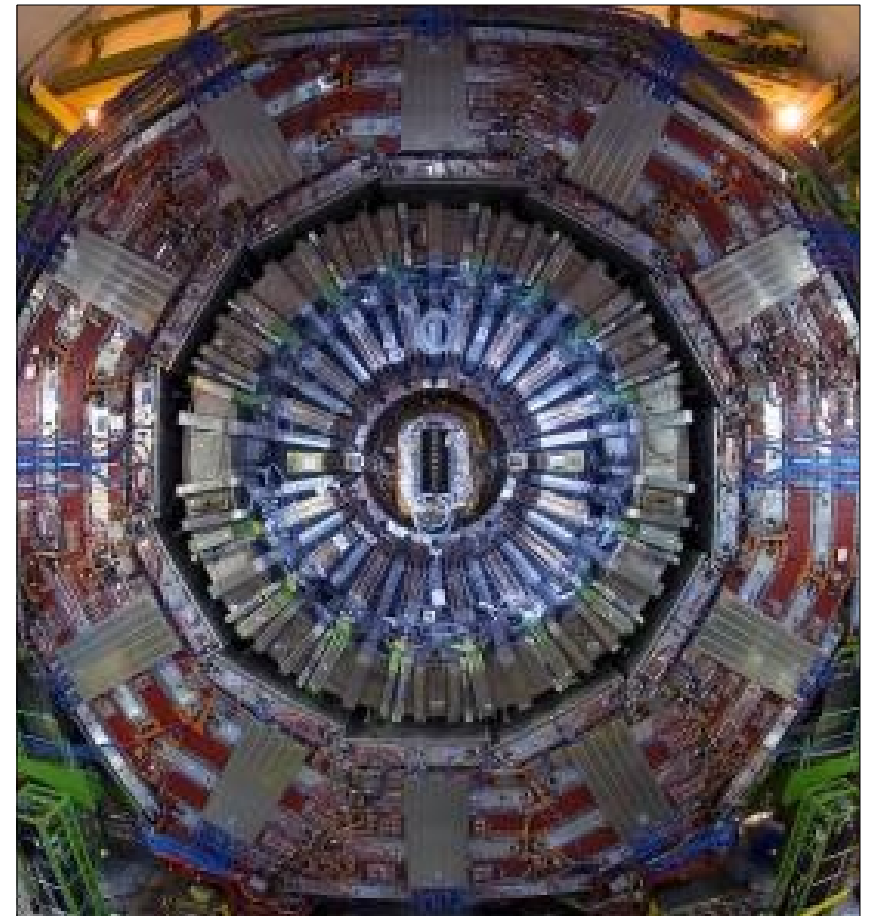
HTCONDOR

SEE, I'LL TALK ABOUT IT!

We still run a fairly substantial HTCondor resource

Serving CMS, opportunistic access for many OSG VOs, and several Purdue researchers

Out of 240M hours delivered by our center last year, ~31M were “High Throughput”



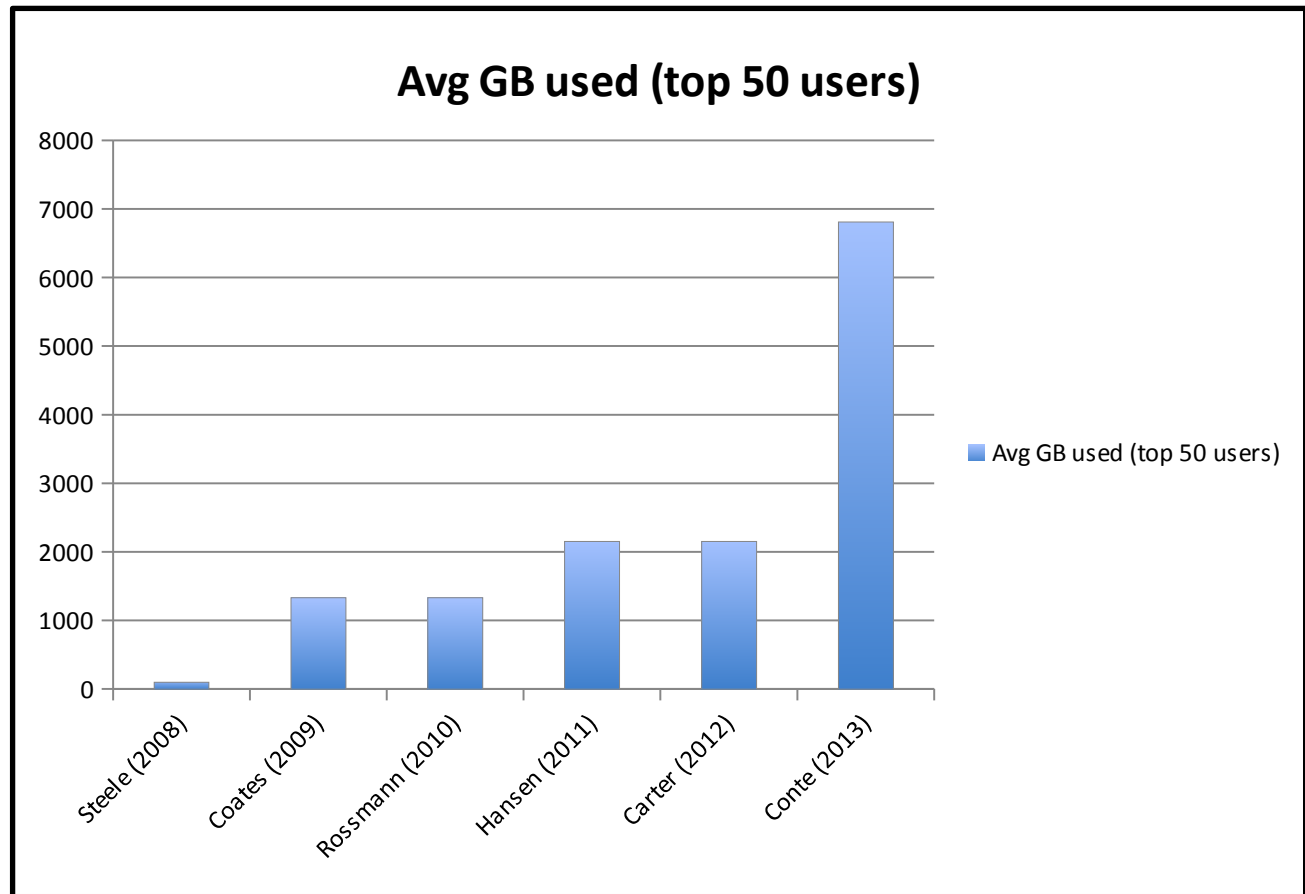


DATA

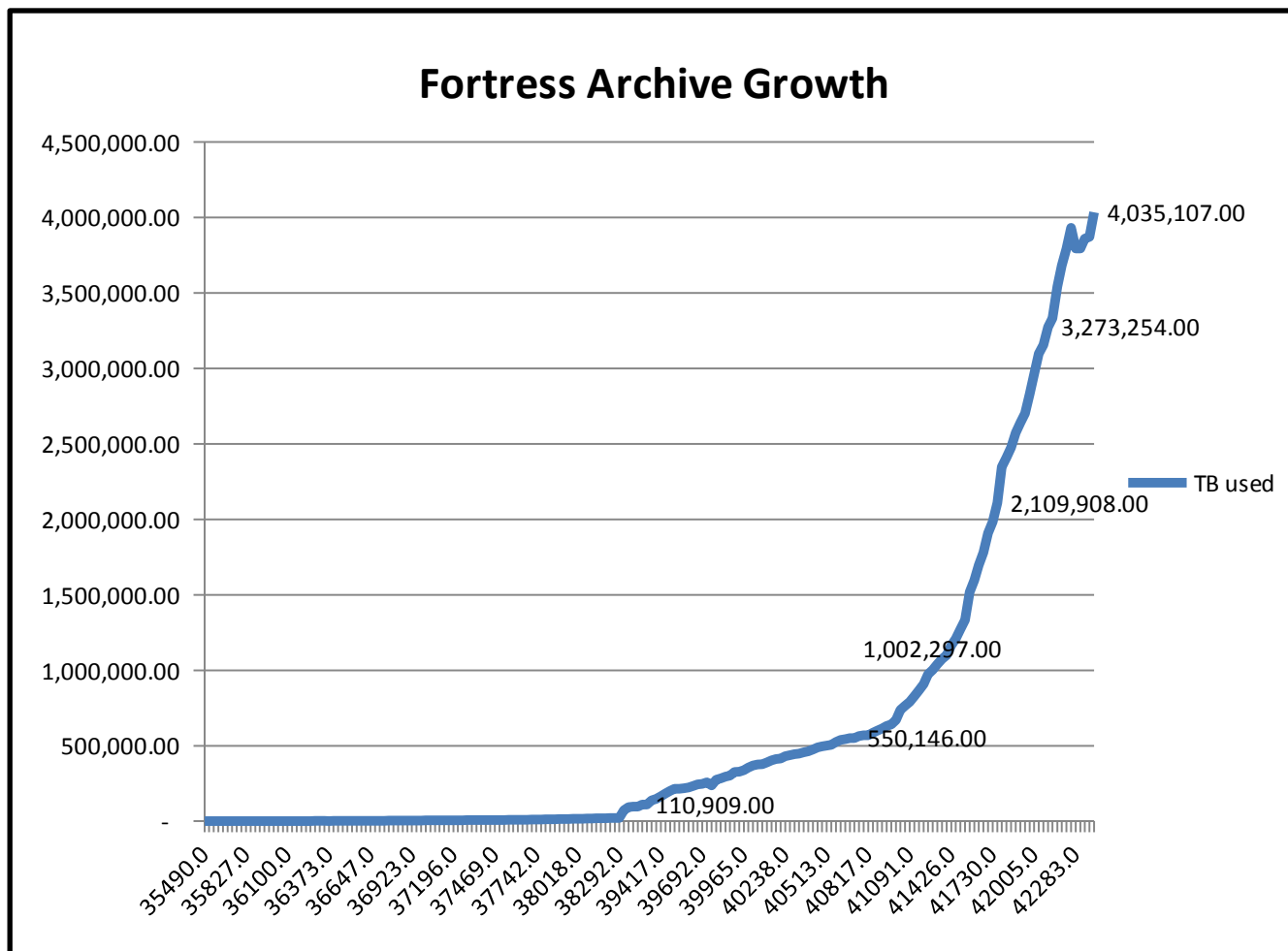
STORAGE

INFRASTRUCTURE FOR RESEARCH DATA

Scratch needs are climbing

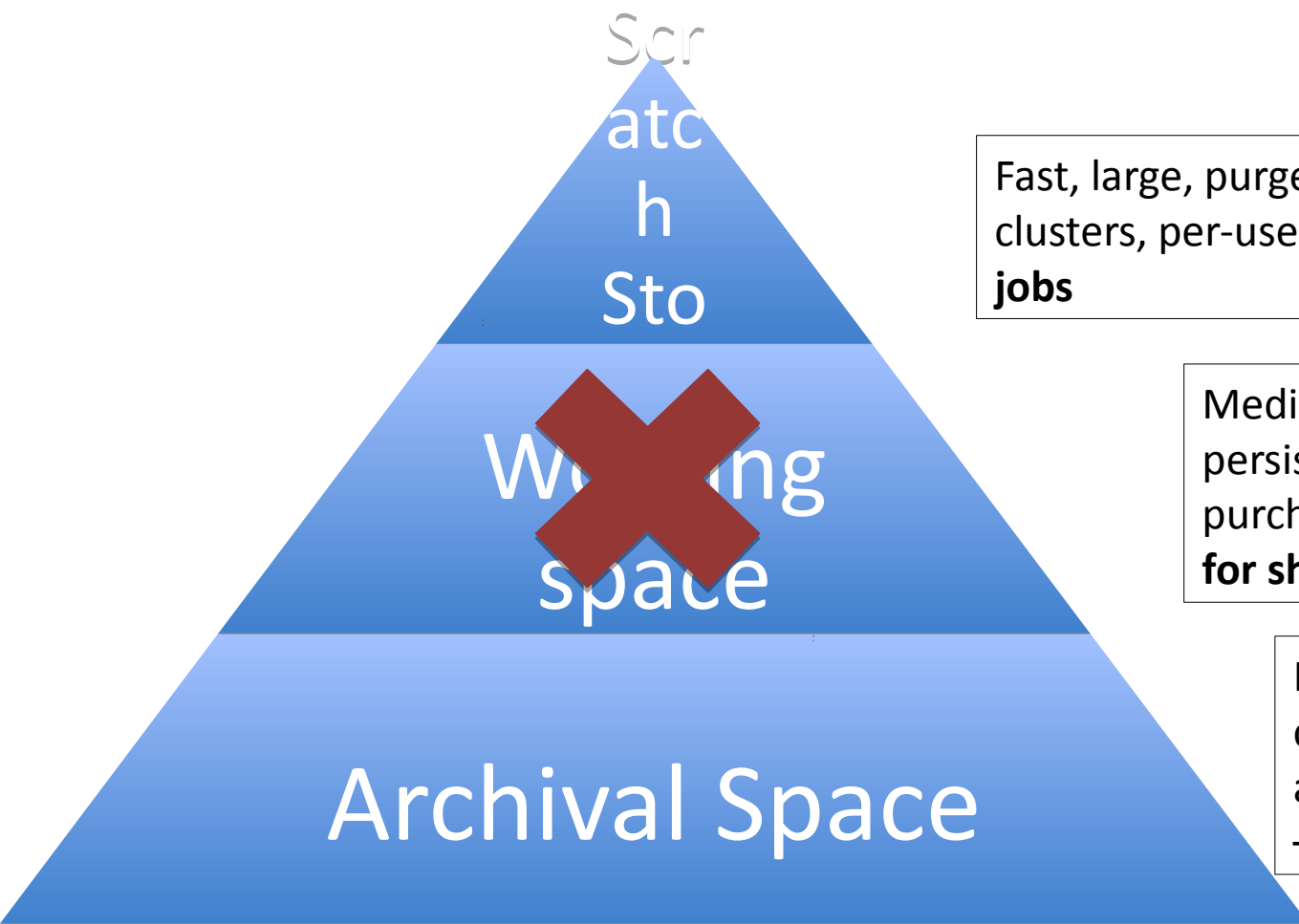


Archive usage is skyrocketing



HPC STORAGE

TIERS OF STORAGE



Fast, large, purged, coupled with clusters, per-user – **for running jobs**

Medium speed, large, persistent, data protected, purchased, per research lab – **for shared data and apps**

High-speed, infinite capacity, highly-protected, available to all researchers – **for permanent storage**

**Working with researchers across campus,
we encounter many different data
solutions..**

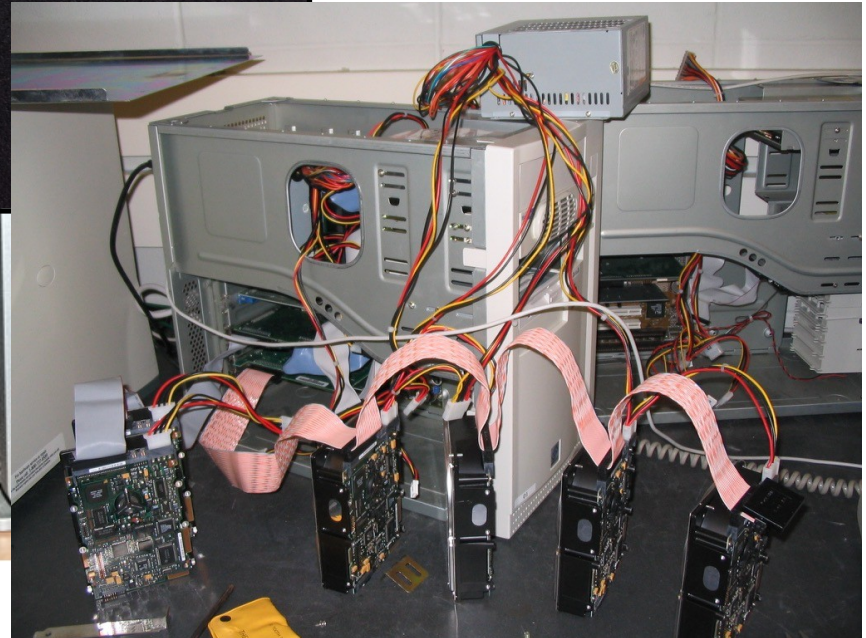
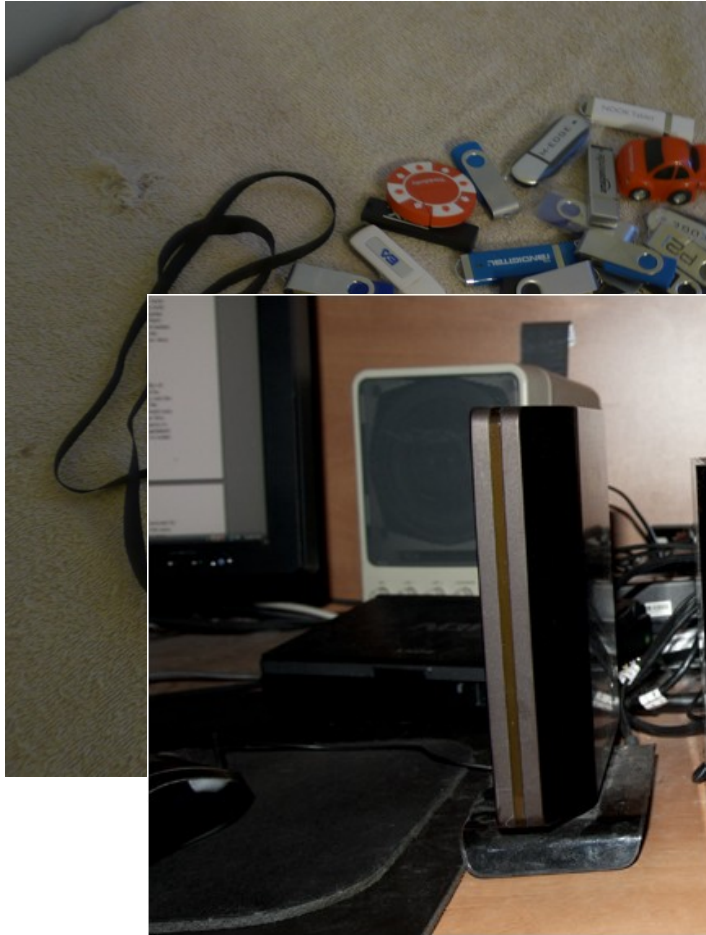
**From something at the
department/workgroup level:**



RESEARCH DATA ACROSS CAMPUS

CURRENT STATE

To This



RESEARCH DATA ACROSS CAMPUS

CURRENT STATE

And This



Many central storage options have not met all the needs that researchers care about

- Departmental or lab resources are islands and not accessible from HPC clusters.
- Most are individually-oriented, rather than built around the notion of a research lab.
- Scratch filesystems are *also* limited in scope to a single HPC system
- Some are not performant enough for research use
- Sometimes nothing is available for sale, despite faculty having funds

In the past, we've heard lots of common requests:

- I need more space than I can get in scratch
- Where can I install applications for my entire research lab?
- I'm **actively working** on that data/software in scratch:
 - I have to go to great lengths to keep it from being purged.
 - I shouldn't have to pull it from Fortress over and over
- Can I get a UNIX group created for my students and I?
- Is there storage that I can get to on **all** the clusters I use?
- I have funding to spend on storage – what do you have to sell?
- I need storage for my instrument to write data into
- My student has the only copy of my research data in his home directory, and he graduated/went off the grid!

RESEARCH STORAGE GAPS

SOME SOLUTIONS

We've addressed some of these with improving scratch:

- Everybody automatically gets access to HPSS Archive for permanent data storage
- Very large per-user scratch quotas
- More friendly purge policy – based on the *use* of data, rather than when it was created.

DEPOT

WHY A DEPOT?

As a transport hub: a place where large amounts of cargo are stored, loaded, unloaded, and moved from place to place.



- It's important to think of Depot as a “data service” – not “storage”
- It is not enough to just provide infrastructure
 - “Here’s a mountpoint, have fun”
- Our goal: enabling the frictionless use and movement of data
 - *Instrument -> Depot -> Scratch -> Fortress -> Collaborators -> and back*
 - Continue to improve access to non-UNIX users

Purdue researchers can purchase storage!

An affordable storage service for research to address many common requests:

- 100G available at no charge to research groups
- Mounted on all clusters and exported via CIFS to labs
- *Not scratch*: Backed up via snapshots, with DR coverage
- Data in Depot is owned by faculty member!
- Sharing ability – Globus, CIFS, and WWW
- Maintain group-wide copies of application software or shared data


A TOOL-FRIENDLY STRUCTURE

AS A STARTING POINT


...with POSIX ACLs to overcome group permission and umask challenges!

SELF-SERVICE


MANAGE YOUR OWN ACCESS

Managers 

	◀◀	Unix Groups			
			-data	-apps	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	

Group Usage Reporting Viewers 

	◀◀	Unix Groups			
			-data	-apps	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Members 

	◀◀	Unix Groups			
			-data	-apps	
		<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

Well received!

- In just over a year, over 260 research groups are participating.
 - *Many are not HPC users!*
- .75 PB in use since November 2014
- A research group purchasing space has purchased, on average, 8.6TB.

A TESTIMONIAL

SATISFIED FACULTY

"It's really, really useful. It's actually one of the best things about this setup. In the past, with other HPC systems I've used, moving all the data around that these models output has always been a major pain.

One of these high resolution simulations can produce terabytes of output and you've got to have somewhere you can put that in order to analyze it. The Research Data Depot is a place where I can put a large amount of data, I can mount it on local machines and access it.

The Research Data Depot [is] large like the scratch but it's disk based and fast unlike the tape archive, so it's got the best of both world's, I think, and it's backed up! I think it's great. I use it a lot. I have 25 terabytes right now and I'd like to get more."

Daniel Dawson, Assistant Professor of Earth,
Atmospheric, and Planetary Science

THE TECHNOLOGY

WHAT DID WE GET?

Approximately 2.25 PB of IBM GPFS

Hardware provided by a pair of Data Direct Networks SFA12k arrays, one in each of two datacenters

160 Gb/sec to each datacenter

5x Dell R620 servers in each datacenter



DESIGN TARGETS

WHAT DO WE NEED TO DO?

The Research Data Depot Can do:

Depot Requirements	Previous solutions
At least 1 PB usable capacity	>1 PB
40 GB/sec throughput	5 GB/sec
< 3ms average latency, < 20 ms maximum latency	Variable
100k IOPS sustained	55k
300 MB/sec min client speed	200 MB/sec max
Support 3000 simultaneous clients	Yes
Filesystem snapshots	Yes
Multi-site replication	No
Expandable to 10 PB	Yes
Fully POSIX compliant, including parallel I/O	No

GROWTH AREAS

GROWTH BY COLLEGE

College	2015	New Groups	2014	Growth rate 2014-2015
Liberal Arts	7	6	1	600%
Education	2	1	1	100%
HHS	22	8	14	57%
Agriculture	74	26	48	54%
Science (bio)	21	6	15	40%
Pharmacy	6	1	5	20%
Engineering	188	27	161	17%
Science (non-bio)	200	16	184	9%
Technology	14	1	13	8%
Management	20	0	20	0%
Vet School	1	1	0	
	555	93	462	20%

Life Science!



RELATED

CHALLENGES

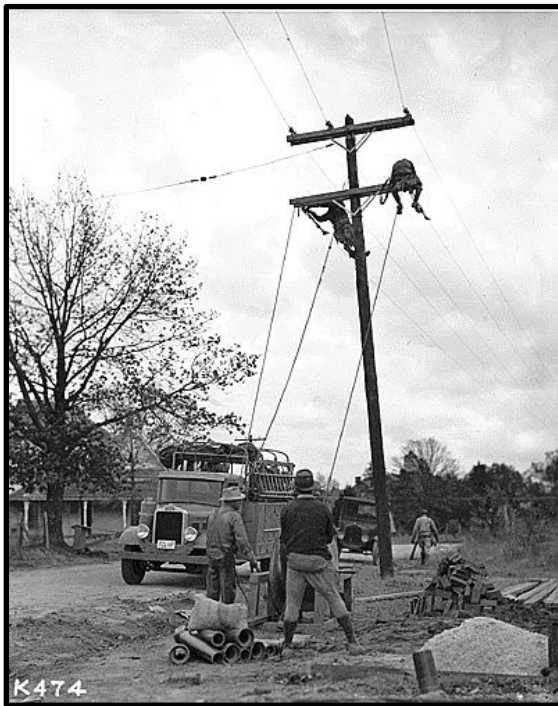
BEYOND THE DATACENTER

IMPROVED NETWORKING

WHAT DOES THIS MEAN FOR RESEARCHERS?

As science gets more data-intensive – researchers require increasing amounts of bandwidth (GigE on up)

The last mile to their labs will be the key!



IMPROVED NETWORKING

INSTRUMENTS

Instruments are getting cheaper, more common, and generate more data.

Researchers need high-speed (10Gb+) connections for labs and instruments to move data into clusters, Research Data Depot, and research WAN connections.



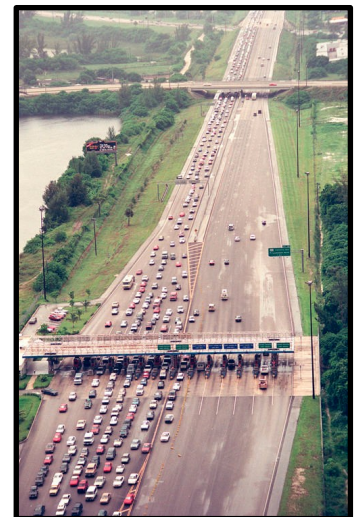
FIREWALLS

SAFETY OR PERFORMANCE

The situation on campus is mixed, but researchers shouldn't have to choose.

Watch for fast GigE links being throttled by 100Mb firewalls!

We need to get researchers from DIY firewalls into the enterprise-grade one.



Data transferred in the last year:

Volume: nearly 300 TB transferred!

Avg 23 TB, 50 unique users each month



Killer feature: sharing!

FUTURE WORK

- Need staff with expertise in the mechanics of working with data
- Work with Libraries to integrate with institutional repository
- How to centrally fund the storage?

PURDUE UNIVERSITY | Purdue University Research Repository | PURR

Home Resources Projects Get Started Contact Us

Publish Datasets with DOIs

Use PURR to publish datasets with Digital Object Identifiers (DOIs) that make it easier for people to cite your data and give you credit. Purdue is a founding member of DataCite, the international agency that registers DOIs for data.

[Learn More](#)

Start Your Research Project

- Create a Data Management Plan: Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP Resources can help you to clear up any confusion. [Get Started](#)
- Upload Research Data to Your Project: Create a project to upload and share your data with collaborators using our step-by-step form to guide you through the process. Invite collaborators from other institutions to join your project. [Create a Project](#)
- Publish your Dataset: Package, describe, and publish your dataset with a DataCite DOI. Publishing will ensure your dataset is citable, reusable, and archived for the long term. [See Published Datasets](#)

Featured Dataset

Evaluation of Function Predictions for Bloodclotting Proteins
By Anita K. Khan, Stephanie Chikan, Catherine Paxon, Devaki Ghosh
Purdue University, University of Florida, A*STAR

Supplemental datasets used for evaluation of function prediction for managing proteins.

Do you have a question?

Ask a librarian

Send us your email question:

THANK YOU!

QUESTIONS?

Questions?

Contact Us:

psmith@purdue.edu

@PurdueRCAC

<http://www.facebook.com/PurdueRCAC>